

Do patient-reported outcomes offer a more sensitive method for comparing the outcomes of consultants than mortality? A multilevel analysis of routine data

Mira Varaganam, Andrew Hutchings, Nick Black

Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, London, UK

Correspondence to

Dr Nick Black, Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK; nick.black@lshtm.ac.uk

Received 5 September 2014
Revised 14 October 2014
Accepted 15 October 2014
Published Online First
3 November 2014

ABSTRACT

Background Patient-reported outcome measures (PROM) might be better for comparing consultant surgeons' outcomes than mortality.

Objectives To describe variation in outcomes between consultants, compare the number of outlying consultants according to different measures, explore the effect that the hospital in which a consultant works has on their outcomes and determine the scope for improving outcomes by reducing variation between consultants.

Method Consultants performing hip replacement (n=948), knee replacement (1130) and hernia repair (974) in National Health Service hospitals in England in 2009–2012; disease-specific and generic PROMs and complications; fixed-effects and multilevel models to assess consultant outcomes, were all compared. Influence of patient factors and hospital factors was assessed.

Results Fixed-effects models showed that most consultants are 'as or better than expected'. However, unlike with mortality, some consultants are more than three SDs 'worse than expected' according to disease-specific PROMs (2.4% for hip and 1.2% for knee replacement), generic PROMs (1.2% and 1.0%) and incidence of complications (1.8% and 0.8%). The proportion of consultants worse than expected is less with random-effects models. Controlling for hospital factors reduced the proportion further. After controlling for known patient characteristics, consultants and hospitals contribute little towards variation in patient outcomes.

Conclusions PROMs offer a more appropriate and sensitive method for comparing consultants' outcomes. The influence of hospitals must be considered to ensure comparisons are meaningful. Improvements will be achieved by shifting the distribution of consultants rather than by reducing variation between them.

INTRODUCTION

In 2013, public reporting of the outcomes of individual consultant surgeons in England was introduced.¹ Information on outcomes was mostly limited to post-operative mortality, the measure already in use in cardiac surgery in the UK² and the USA.³ The aims were to stimulate improvement in the quality of care among those found to be 'worse than expected' by encouraging them and their managers to redesign their services and by allowing patients to make informed choices when selecting a consultant. However, despite such worthy intentions, the approach was limited for three reasons. First, as the outcome is rare, it is unlikely that a poorly performing consultant would be detected.⁴ For example, the 90-day mortality following hip replacement in 2011 was 0.3%,⁵ and for knee replacement averaged 0.4% between 2003 and 2013.⁶ These low proportions meant that a consultant's work over the preceding 10 years had to be considered, undermining the relevance of the assessment for a patient wanting information on the current outcomes a consultant was achieving. As could have been predicted, the approach did not identify any consultants as worse than expected (defined as more than three SDs below the mean)⁶ with the risk that the whole exercise might have engendered a false sense of complacency.

The second limitation was that the analysis treated the outcomes of consultants as fixed effects, so it was not possible to investigate the hierarchical nature of the data and consider that variation in consultants' outcomes might have been due



To cite: Varaganam M, Hutchings A, Black N. *BMJ Qual Saf* 2015;**24**:195–202.

to variation in hospitals' performance. And third, postoperative mortality is a measure of safety and provides no indication of how effective surgery is in improving patients' health status and health-related quality of life (HRQL). Given that most elective surgery is carried out to reduce patients' symptoms, improve functioning and enhance quality of life, these are the appropriate outcome measures to use.

Whereas in the past, routine data on these aspects were not available, accurate data are starting to be collected in England through the routine use of patient-reported outcome measures (PROM). These have the advantages of providing a continuous variable (rather than the dichotomy of dead or alive) and of ensuring the outcomes of all patients contribute to the determination of a consultant's outcome. Additionally, in England, patients also report on the occurrence of common complications, providing a measure of safety that overcomes some of the statistical problems encountered with the rarity of postoperative death.

Thus, the availability of the National PROMs Programme in England since 2009 provides a means of investigating the potential of PROMs for comparing consultants' outcomes.⁷ To date, use of such data has been restricted to comparing hospitals.⁸ This is despite evidence that patients wish to be able to choose their surgeon rather than just their hospital,⁹ plus the wish of politicians for public information on the outcomes of individual surgeons.

Our aim was to explore the use of PROMs as an indicator of consultants' outcomes. Our objectives were to: describe the extent of variation in outcomes between consultants; compare the number of outlying consultants according to different outcomes; explore the effect that the hospital in which a consultant works has on their outcomes and determine the scope for improving outcomes by reducing variation between consultants.

METHODS

Data

We conducted a secondary analysis of existing routine data. All patients undergoing one of three procedures (hip replacement, knee replacement, groin hernia repair) between April 2009 and March 2012 were invited to complete a preoperative questionnaire which included PROMs. (The data for varicose vein surgery was not considered to be of sufficient quality to use in this analysis). Follow-up questionnaires mailed to patients 3 months (hernia repair) or 6 months (hip and knee replacement) after surgery included the same PROMs and a question about the occurrence of four common complications (wound problems; urinary problems; allergy or reaction to drug; bleeding).

Completed questionnaires were linked by the Health and Social Care Information Centre, to the patient's episode in the hospital administrative database—Hospital Episode Statistics (HES). This provided

information on the patient's ethnicity and enabled us to assign their socioeconomic status (derived from their postcode and based on the Index of Multiple Deprivation (IMD) using 2007 rankings).¹⁰ Additionally, HES identified the patient's consultant and the hospital where the operation was performed.

The effectiveness of hip and knee replacement was assessed using a disease-specific PROM—the Oxford Hip Score (OHS)¹¹ and the Oxford Knee Score (OKS).¹² Both include 12 items, each scored from 0 to 4 and summated to provide an overall score of between 0 (severe symptoms and disability) and 48 (no problem). There was no disease-specific PROM available for hernia repair. All three procedures were also assessed using a generic PROM, the EQ-5D-3L which assesses five dimensions of health (mobility, self-care, daily activities, pain and anxiety/depression).¹³ Each question has three levels of response, and answers are transformed into an HRQL score using material from the UK-Time Trade-Off value set.¹⁴ Scores range from −0.59 (worse than death), through 0.00 (dead) to 1.00 (perfect health). Finally, the safety of all three procedures was assessed as the proportion of patients reporting at least one of four common complications. Analysing each complication separately would have been less likely to detect an outlier.

Inclusion criteria for hospitals and consultants

Analyses were based on National Health Service (NHS)-funded patients treated by NHS providers (hospitals and treatment centres). Independent (private) hospitals and independent sector treatment centres were excluded. NHS providers were included if over the 3 years there were PROMs data for 40 or more patients. The consultants working in the included providers were included if data were available for at least 10 patients who they were clinically responsible for (though they may not have been the operating surgeon). The adoption of these thresholds represented a pragmatic balance between needing to exclude those for whom the amount of data was too small to be representative while excluding as few as possible. The numbers of hospitals and consultants who met the inclusion criteria are shown in [table 1](#).

Extent of variation in outcomes between consultants: fixed-effects model

We adjusted for patients' age, sex, socioeconomic status (as quintiles of the IMD), self-reported comorbidities (heart disease; hypertension; stroke; claudication; lung disease; diabetes; kidney disease; neurological diseases; liver disease; cancer; depression), having had previous surgery on the same hip or knee (for those operations), and preoperative PROM scores. The last of these is known to be the most predictive factor.¹⁵ We used multivariable linear regression for the continuous measures and logistic

Table 1 Number of hospitals and consultants meeting inclusion criteria for each procedure; and number of patients completing preoperative (Pre-op) and postoperative (Post-op) patient-reported outcome measures (PROM) (2009–2012)

	Hip replacement	Knee replacement	Hernia repair
Hospitals (PROMs for 40 or more patients)	183	188	197
Consultants (PROMs for 10 or more patients)	948	1130	974
Patients (Pre-op and post-op PROMs and HES linkage)	65 465	68 107	38 965

HES, Hospital Episode Statistics.

regression for the binary outcome measure of complications. Preoperative PROM scores were modelled as linear plus quadratic terms to allow for non-linearity in their relationship with the outcome.

A measure of each consultant's case-mix adjusted outcome was then estimated: for the continuous measure, this was the overall mean score plus the consultant's mean residual score (actual minus predicted mean score); for the binary outcome measures, this was the overall proportion multiplied by the provider's ratio of their actual to their predicted proportion.¹⁶ Outliers were defined as more than three SDs from the mean score. Consultants were compared for three outcomes: case-mix adjusted postoperative PROM score (disease-specific and generic) and the proportion of patients reporting a complication.

Effect of the hospital on consultants' outcomes: random-effects model

Multilevel modelling, adjusting for hospital and patient differences, is akin to comparing consultants as if they were working in the same hospital and operating on the same type of patient. Hence, any difference in the outcome can be fairly attributed to the consultant.

To consider the impact of a hospital on patients' outcomes independent of that of a consultant, we used a random-effects model which has better reliability than a fixed-effects model to detect true quality differences.¹⁷ Hospital factors that might have an impact include staffing levels, nursing skills, pharmaceutical use and length of stay. However, no such covariates were fitted at the hospital or consultant levels in the present study as no data were available.

To enhance the ability of the analysis to detect any impact hospitals might have on consultants' outcomes (given that wider CIs are to be expected with random-effects models), consultants were deemed to be outliers if they were more than two SDs (rather than three SDs) from the mean. First, the proportions of outliers, controlling only for patient variation, were derived and compared with the proportions derived from the fixed-effects model. Then, hospital variation was introduced and controlled for. The CI for consultant performance was calculated by adding the consultant level residual to ± 2 SEs of the residual.

To consider the relationship between consultant and hospital outcome, estimates for consultant outcome

were derived from a two-level model with patients and consultants. Estimates for hospital outcome were derived from a two-level model with patients and hospitals. Estimates for consultant outcome after controlling for hospital effects were derived from a three-level model with patients, consultants and hospitals.

Assess the determinants of residual variation between consultants

To determine the contributions of consultants and hospitals to variation in patients' outcomes after adjusting for case-mix, progressively complex multi-level models (using iterative generalised least square estimation) with disease-specific and generic PROMs scores as the outcome variables were run. The likelihood ratio test was performed to determine the best model that fitted the national data. One model had a random intercept for consultants; a second model had random intercepts for hospitals and consultants. The best fitting model showed variation at the level of the hospital and consultant for hip and knee replacement and at the level of hospital for hernia repair.

For patient-reported complications, progressively complex logit multilevel models (using Markov's Chain Monte Carlo Method) were performed with the binary outcome. Model fit was assessed using the deviance information criterion. The best fitting model for all three procedures was one with a random intercept at the level of hospital and consultant. Partitioning of variance was calculated using the linear threshold representation where the patient-level residuals follow a logistic distribution with a constant variance of 3.29.

All analyses were performed using STATA V.11.¹⁸ Multilevel modelling was performed using RunMLwin (University of Bristol) which allows RunMLWin to run within STATA.¹⁹

RESULTS

Comparison of consultants' outcomes

The fixed-effects model revealed the extent to which consultants' outcomes varied (table 2). The proportions of 'good' outliers (more than 3 SDs above average) and 'bad' outliers (more than 3 SDs below average) according to the disease-specific PROMs were both greater than would be expected by chance (0.1% above and 0.1% below average) for hip and knee replacement. For example, of the 948

Table 2 Proportions of consultants with mean outcomes more than three SDs different than expected

Procedure and outcome	Number (%) of consultants		
	Better than expected	As expected	Worse than expected
Hip replacement			
Oxford Hip Score	10 (1.1)	915 (96.5)	23 (2.4)
EQ-5D	7 (0.7)	930 (98.1)	11 (1.2)
Complications	8 (0.8)	923 (97.4)	17 (1.8)
Knee replacement			
Oxford Knee Score	12 (1.1)	1104 (97.7)	14 (1.2)
EQ-5D	6 (0.5)	1113 (98.5)	11 (1.0)
Complications	9 (0.8)	1114 (98.4)	7 (0.6)
Hernia repair			
EQ-5D	0 (0.0)	972 (99.8)	2 (0.2)
Complications	2 (0.2)	971 (99.7)	1 (0.2)

consultants performing hip replacement, 10 (1.1%) had adjusted postoperative OHS that were more than three SDs 'better than expected', and for 23 (2.4%) were worse than expected. There were fewer outliers detected using the EQ-5D, though still more than would be expected by chance for hip and knee replacements. For hernia repair, there were no more outliers than expected by chance.

Effect of the hospital on consultants' outcomes

The impact of using a random-effects (controlling only for patient variation) model rather than fixed-effects model was, as expected, a reduction in the proportion of consultants deemed to be outliers. For this reason, we defined outliers as two SDs away from the mean, to increase sensitivity (table 3). For hip replacement, the proportion (better and worse than expected combined)

Table 3 Proportions of consultants with mean outcomes more than two SDs different than average

Procedure	Number (%) of consultants		
	Better than expected	As expected	Worse than expected
Hip replacement			
Disease-specific PROM (OHS)	53 (5.6)	820 (86.5)	75 (7.9)
▶ Fixed-effect model			
▶ Random-effects model			
Controlling for patient variation	52 (5.5)	856 (90.3)	40 (4.2)
Controlling for patient and hospital variation	7 (0.7)	926 (97.7)	15 (1.6)
EQ-5D	50 (5.3)	848 (89.4)	50 (5.3)
▶ Fixed-effects model			
▶ Random-effects model			
Controlling for patient variation	34 (3.6)	900 (94.9)	14 (1.5)
Controlling for patient and hospital variation	5 (0.5)	940 (99.2)	3 (0.3)
Knee replacement			
Disease-specific PROM (OKS)	47 (4.2)	1024 (90.6)	59 (5.2)
▶ Fixed-effects model			
▶ Random-effects model			
Controlling for patient variation	31 (2.7)	1071 (94.8)	28 (2.5)
Controlling for patient and hospital variation	10 (0.9)	1115 (98.7)	5 (0.4)
EQ-5D	35 (3.1)	1055 (93.4)	40 (3.5)
▶ Fixed-effects model			
▶ Random-effects model			
Controlling for patient variation	11 (1.0)	1112 (98.4)	7 (0.6)
Controlling for patient and hospital variation	2 (0.2)	1126 (99.6)	2 (0.2)
Hernia repair			
EQ-5D	4 (0.4)	944 (96.9)	26 (2.7)
▶ Fixed-effects model			
▶ Random-effects model			
Controlling for patient variation	0 (0.0)	974 (100%)	0 (0.0)
Controlling for patient and hospital variation	0 (0.0)	974 (100%)	0 (0.0)

OHS, Oxford Hip Score; OKS, Oxford Knee Score; PROM, patient-reported outcome measure.

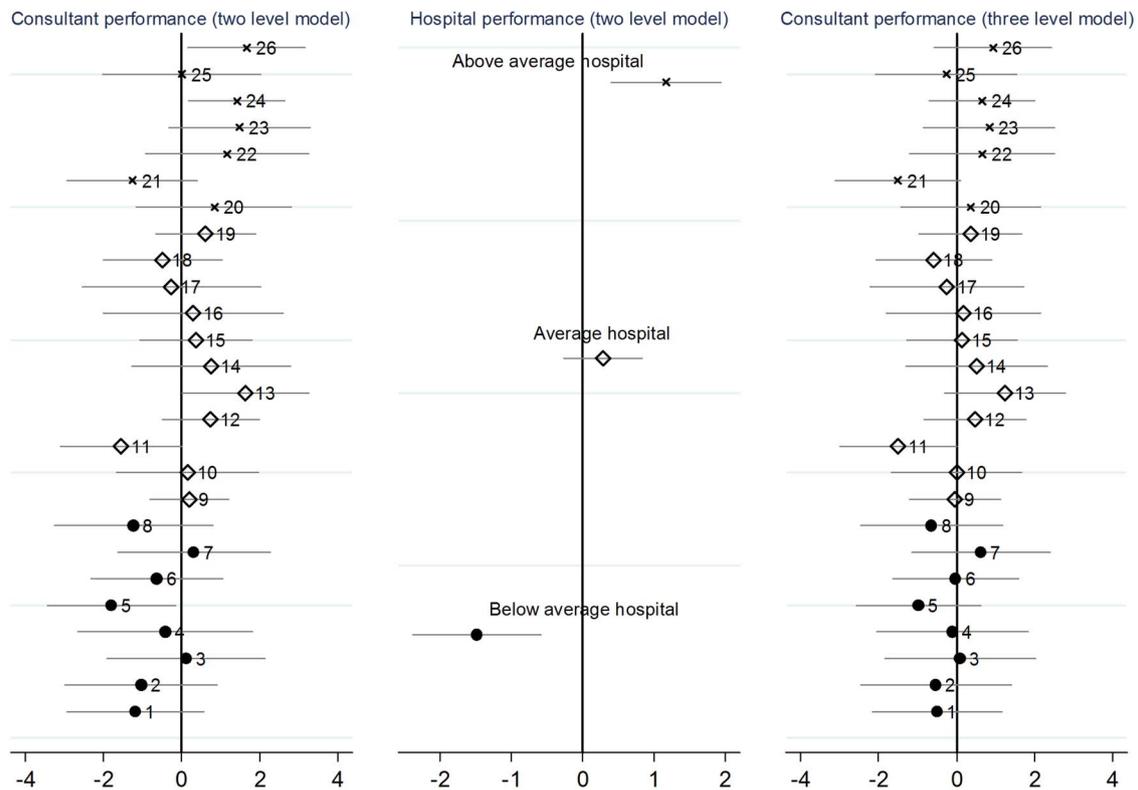


Figure 1 Influence of hospital factors on consultants' outcomes for hip replacement (postoperative Oxford Hip Score): examples of an above average hospital (with seven consultants); average hospital (11 consultants) and below average hospital (eight consultants).

according to the disease-specific PROM was 9.7% with the random-effects model compared with 13.5% with the fixed-effects model, and with the EQ-5D it was 5.1% compared with 10.6% (table 3). A similar pattern was seen with knee replacement and hernia repair. For all procedures there were still more 'outliers' than would be expected by chance.

Controlling for hospital variation led to fewer consultants being deemed outliers. For hip replacement, only 22 (2.3%) consultants were still outliers according to the disease-specific PROM (fewer than the 5% expected by chance) and 8 (0.8%) according to the EQ-5D. A similar pattern was seen with knee replacement. For hernia repair, given that no consultant was more than two SDs from the mean after controlling for patient variation, controlling for hospital could have no impact.

One way of observing the influence of a hospital on a consultant's outcomes is to consider three situations: hospitals categorised as below average, average and above average. In figure 1, consultant and hospital outcomes are indicated by the residual in the random-effects models, that is, the extent to which outcome is explained by the consultant or hospital. The central panel shows the outcome of hip replacement according to case-mix adjusted postoperative disease-specific PROM (OHS) for three contrasting hospitals. These data reflect the influence both of consultants and hospital factors. The left-hand panel shows the outcomes for the consultants working in

each hospital: consultants 1–8 work in a below average hospital (●), 10–19 work in an average hospital (◇) and 20–26 work in an above average hospital (×). The right-hand panel shows the consultants' outcomes after the effect of the hospital is taken into account. For example, in the below average hospital, none of the eight consultants are outliers according to the three-level model (right-hand panel), suggesting that the poor patient outcomes arise from hospital factors rather than below average consultants.

By contrast, consultant 13 who is above average (left-hand panel) is working in an average hospital. After adjusting for hospital effects, he appears to be only average (right-hand panel) suggesting his outcome is dragged down by hospital factors. The converse can also occur: consultant 5 is below average (left-hand panel), but after adjusting for hospital effects he appears to be only average, suggesting that his below average outcome was partly the result of hospital rather than personal factors.

Of those considered better than expected in the three-level model, some may be missed and appear 'as expected' with the two-level model. For example, this was true for one of the seven consultants better than expected for hip replacement using OHS in the three-level model.

Determinants of residual variation between consultants

To address the scope for improvement in outcomes that might be possible, the proportion of the residual

variation for the three procedures using the disease-specific PROM for hip and knee replacement (OHS and OKS, respectively), and the EQ-5D score for hernia repair were considered (table 4). Almost all the residual variation (eg, 97% for hip replacement) is attributable to patient characteristics that are either not adequately accounted for by the risk adjustment model or have not been measured and included in the model. Very little of the variation can be attributed to characteristics of the consultant or hospital (eg, for hip replacement 1.6% and 1.0%, respectively). The same was true for all three procedures and for the rates of complications.

DISCUSSION

Main findings

Unlike comparisons based on 90-day postoperative mortality, use of patients' reports of postoperative symptoms, functional status, quality of life and incidence of complications reveals significant variation between consultants' outcomes. When comparing the three available outcome measures, the proportion of consultants deemed worse than expected is higher according to disease-specific PROMs than when generic PROMs or the incidence of complications are considered.

When random-effects models are used instead of fixed effects, the number of worse than expected consultants decreases (eg, 40 instead of 75 for hip replacement according to the disease-specific PROM), but is still greater than would be expected by chance. The smaller number is not surprising as random-effects models are conservative in identifying outliers compared with fixed-effects models. This has previously been shown in comparisons of hospital mortality following cardiac surgery²⁰ and acute myocardial infarction.²¹ This raises the issue of which type of model is more appropriate for quality improvement purposes. Fixed-effects approaches have greater sensitivity (ie, more likely to detect true outliers) but lower specificity (ie, more likely to incorrectly classify consultants as outliers). Random-effects approaches have

lower sensitivity but higher specificity; so the choice depends on whether sensitivity or specificity is of greater policy importance.

When variation at the hospital level is taken into account, in addition to patient variation, the number of outlier consultants decreases further and is no longer more than might be expected by chance, for example, only 15 of the 40 consultants remain worse than expected for hip replacement, according to the disease-specific PROM, and only five out of 28 for knee replacement. This suggests that the majority of consultants who appear to be worse than expected are being disadvantaged by practicing in a hospital with poor outcomes. Efforts at quality improvement directed at consultants in such situations are, therefore, unlikely to improve outcomes.

The lack of scope for improvement was also evident when explanations for residual variation were explored. This showed that apart from that due to chance, characteristics of consultants and hospitals explained very little for all three procedures.

Limitations of the study

There are several potential limitations to consider. First, despite including a large number of hospitals, consultants and patients, we confined the sample to NHS-funded patients who underwent surgery in NHS hospitals and treatment centres. We excluded NHS-funded patients treated by private providers and privately funded patients. Excluded patients may have been at lower risk of a poor outcome: those treated in independent sector treatment centres are deliberately selected on this basis, and those paying privately are more likely to be of higher socioeconomic status enjoying access to better postoperative rehabilitation. However, these exclusions will only have narrowed the range of data values and not affected the internal validity of the analyses.

Second, we excluded hospitals if there were fewer than 40 patients with completed PROMs data linked to HES (resulting in excluding 52 (22%) hospitals for hip replacement; 48 (20%) for knee replacement;

Table 4 Variance (SE) and proportion of variance in consultants' outcomes attributable to the hospital, consultant and patient

	Hip replacement		Knee replacement		Hernia repair	
	Variance (SE)	Proportion of variance	Variance (SE)	Proportion of variance	Variance (SE)	Proportion of variance
PROM*						
Hospital	0.61 (0.12)	0.8%	0.57 (0.10)	0.7%	0.00006 (0.00002)	0.1%
Consultant	1.16 (0.12)	1.6%	0.79 (0.10)	1.0%	0.00	0.0%
Patient	71.89 (0.40)	97.0%	80.50 (0.44)	98.0%	0.03 (0.0002)	99.9%
Complication						
Hospital	0.03 (0.006)	0.9%	0.02 (0.003)	0.6%	0.02 (0.006)	0.6%
Consultant	0.02 (0.005)	0.6%	0.02 (0.004)	0.6%	0.03 (0.009)	0.6%

*Oxford Hip Score for hip replacement; Oxford Knee Score for knee replacement; EQ-5D for hernia repair. PROM, patient-reported outcome measure.

97 (33%) for hernia repair). These included low-volume providers (in the case of hernia repair), poor recruitment of patients to the PROMs programme, poor postoperative response rates and failure to link the PROMs data to the appropriate HES episode. Exclusions might have led to a slight underestimation of the proportion of outliers, as those excluded are more likely to have worse outcomes.^{22–24}

Third, within the participating hospitals, not all eligible patients were included as some did not complete a preoperative questionnaire (22% of hip replacements, 19% of knee replacements and 46% of hernia repair patients),²⁵ some did not complete a postoperative questionnaire (15% hip and knee replacement, 27% hernia repair),²⁶ and some could not be linked to their HES record (linkage bias). However, this will not have biased the results, as patient characteristics known to be associated with lower recruitment or response (older, sicker, non-Caucasian and more socially deprived patients) were adjusted for in the analyses. However, the possibility of unknown confounders remains (ie, patients with worse outcomes being less likely to complete postoperative questionnaires) though the extent of any bias is likely to be slight and not have affected the internal validity of the analyses.²⁷

Fourth, as with any risk adjustment model, it does not take every conceivable risk factor into account. For example, patients undergoing hip replacement will not gain as much benefit from surgery if they also have osteoarthritis of their knees or of the other hip joint. Our analysis assumes the prevalence of such patients is evenly distributed across consultants.

Fifth, some consultants work in more than one hospital, whereas the multilevel models assumed a hierarchy of patients nested in consultants nested in hospitals. To explore this, cross-classified models were created, but they did not give a better fit to the data compared with three-level hierarchical models. Hence, we used the latter, more parsimonious, models for our analysis.

Sixth, the measures of outcome used inevitably have limitations in their measurement properties. These are not thought to be serious. Criticisms that the OHS and OKS display a ceiling effect is to ignore the fact that many patients do indeed achieve perfect hip or knee function following surgery, so will be expected to reach the top of the scale. Concerns about the sensitivity of the EQ-5D are also misplaced as scores correlate highly with disease-specific scales.²⁸

Finally, we attributed patients to the consultant who was responsible for their clinical management. In some cases, the operation will have been performed by surgeons working under the consultant. Consultant-level analyses, therefore, reflect the outcomes of all the patients for whom the consultant is responsible and is not restricted to those he or she operated on.

Implications

This is the first attempt to use PROMs data to compare consultants' outcomes on a national scale. All previous reports have been limited to a measure of safety (postoperative mortality) which, given the rarity of death, did not identify any outliers in orthopaedic surgery. The use of PROMs has three advantages over the 90-day mortality rate: it provides information on effectiveness in terms of the principal objectives of reducing disability and increasing HRQL; it is a more sensitive measure, suggesting that there are some outliers, although very small proportions; and it requires data only on clinical practice of 3 years rather than that of 10 years. Thus, PROMs appear to be a better indicator than postoperative mortality for comparing consultants and should be considered as a better alternative.

Second, multilevel models should be used to assess consultants as it is important to distinguish between consultants' and hospitals' contributions to differences in patients' outcomes. Consultants classified as 'worse than expected' might be disadvantaged if they are working in a poorly performing hospital. Controlling for hospital effects may allow a fairer assessment of their outcome though this in some ways ignores the complex interactions between consultants and hospitals.

The third implication arises from the finding that hospital and consultant factors explain very little of the residual variation after adjustment for patient case-mix. Given that most of the residual variation is attributable to chance or to unmeasured or inadequately measured patient factors, this suggests that there is little opportunity to reduce variation between consultants through quality improvement initiatives (such as additional training). If there is scope for improvements in outcome it will be achieved by shifting the whole distribution of consultants rather than by reducing variation between them.

Finally, two notes of caution are needed. (1) These results may not be generalisable to other elective operations, let alone non-elective reasons for admission to hospital. The relative lack of variation in outcomes between consultants seen for these surgical operations may reflect the success of the long-established policy in the NHS to ensure a similar level of satisfactory care across the whole country. (2) Just because these quantitative measures of consultant outcomes failed to show much potential scope for improvement, it does not mean that there are no modifiable factors that might be detected using qualitative methods.²⁹

Acknowledgements We thank Christopher Charlton of University of Bristol for providing support for RunMLWin and three reviewers for helpful comments.

Contributors All three authors contributed to the planning, conduct and reporting of the work described in the article. NB conceived the idea; MV undertook the analyses, supervised by AH; NB drafted the paper and MV and AH commented and contributed to revising it. NB acts as guarantor.

Competing interests None.

Provenance and peer review Not commissioned; internally peer reviewed.

Transparency declaration NB affirms that the manuscript is an honest, accurate and transparent account of the study being reported, that no important aspects of the study have been omitted and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

REFERENCES

- NHS Commissioning Board. Everyone counts: planning for patients 2013/14. 2012. <http://www.commissioningboard.nhs.uk/files/2012/12/everyonecounts-planning.pdf> (accessed 22 May 2014).
- Society for Cardiothoracic Surgery in Great Britain and Ireland. UK surgeons' results. <http://www.scts.org/modules/surgeons/default.aspx> (accessed May 2014).
- Hannan EL. Measuring hospital outcomes: don't make perfect the enemy of good! *J Health Serv Res Policy* 1998;3:67–9.
- Walker K, Neuburger J, Groene O, *et al*. Public reporting of surgeon outcomes: low numbers of procedures lead to false complacency. *Lancet* 2013;382:1674–7.
- Hunt LJ, Ben-Shlomo Y, Clark EM, *et al*. 90-day mortality after 409 096 total hip replacements for osteoarthritis, from the National Joint Registry for England and Wales: a retrospective analysis. *Lancet* 2013;382:1097–104.
- National Joint Registry. Surgeon and hospital profile. <http://njrsurgeonhospitalprofile.org.uk/>
- Guidance on the routine collection of Patient Reported Outcome Measures (PROMs) for the NHS in England*. Department of Health 2009/10. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_092625.pdf (accessed 22 May 2014).
- Neuburger J, Hutchings A, van der Meulen J, *et al*. Using patient reported outcomes (PROs) to compare the performance of providers: does the choice of measure matter? *Med Care* 2013;51:517–23.
- Hildon Z, Allwood D, Black N. Patients' and clinicians' views of comparing the performance of health care providers: a qualitative study. *Health Expect* 2014. doi:10.1111/hex.12037
- Department for Communities and Local Government. *The English indices of deprivation*. London: Communities and Local Government Publications, 2007.
- Dawson J, Fitzpatrick R, Carr A, *et al*. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996;78-B:185–90.
- Dawson J, Fitzpatrick R. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br* 1998;80-B:63–9.
- 199–208. > The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- Dolan P. Modelling valuations for EuroQol health states. *Med Care* 1997;35:1095–108.
- Northgate Information Solutions. PROMs risk adjustment methodology guide for general surgery and orthopaedic procedures. September 2010. <http://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2013/07/proms-ris-adj-meth-sur-orth.pdf> (accessed 25 Sep 2014).
- Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med* 2005;24:1185–202.
- Krell RW, Hozain A, Kao LS, *et al*. Reliability of risk-adjusted outcomes for profiling hospital surgical quality. *JAMA Surg* 2014;149:467–74.
- StataCorp. 2009. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP.
- Leckie G, Charlton C. runmlwin—a program to run the MLwiN multilevel modelling software from within Stata. *J Stat Software* 2013;52:1–40.
- Glance LG, Dick A, Osler TM, *et al*. Impact of changing the statistical methodology on hospital and surgeon ranking: the case of the New York State cardiac surgery report card. *Med Care* 2006;44:311–19.
- Peter C, Alter DA, Tu JV. The use of fixed and random-effects models for classifying hospitals as mortality outlier: a Monte Carlo assessment. *Med Decis Making* 2003;23:526–39.
- Black NA, Johnston A. Volume and outcome in hospital care: evidence, explanations and implications. *Health Serv Manage Res* 1990;3:108–14.
- Chowdhury MM, Dagash H, Pierro A. A systematic review of the impact of volume of surgery and specialization on patient outcome. *Br J Surg* 2007;94:145–61.
- Nordin B, van der Linden W. Volume of procedures and risk of recurrence after repair of groin hernia: national register study. *BMJ* 2008;336:934–7.
- Hutchings A, Neuburger J, van der Meulen J, *et al*. Estimating recruitment rates for routine use of patient reported outcome measures and the impact on provider comparisons. *BMC Health Serv Res* 2014;14:66.
- Hutchings A, Neuburger J, Gross Frie K, *et al*. Factors associated with non-response in routine use of patient reported outcome measures after elective surgery in England. *BMC Health Qual Life Outcomes* 2012;10:34.
- Hutchings A, Gross Frie K, Neuburger J, *et al*. Late response to patient reported outcome questionnaires was associated with worse outcome. *J Clin Epidemiol* 2013; 66:218–25.
- Grosse Frie K, van der Meulen J, Black N. Relationship between patients' reports of complications and symptoms, disability and quality of life after surgery. *Brit J Surg* 2012;99:1156–63.
- Birkmeyer JD, Finks JF, O'Reilly A, *et al*. Surgical skill and complication rates after bariatric surgery. *N Engl J Med* 2013;369:1434–42.

Do patient-reported outcomes offer a more sensitive method for comparing the outcomes of consultants than mortality? A multilevel analysis of routine data

Mira Varaganam, Andrew Hutchings and Nick Black

BMJ Qual Saf 2015 24: 195-202 originally published online November 3, 2014

doi: 10.1136/bmjqs-2014-003551

Updated information and services can be found at:
<http://qualitysafety.bmj.com/content/24/3/195>

These include:

References

This article cites 21 articles, 4 of which you can access for free at:
<http://qualitysafety.bmj.com/content/24/3/195#BIBL>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:
<http://group.bmj.com/subscribe/>